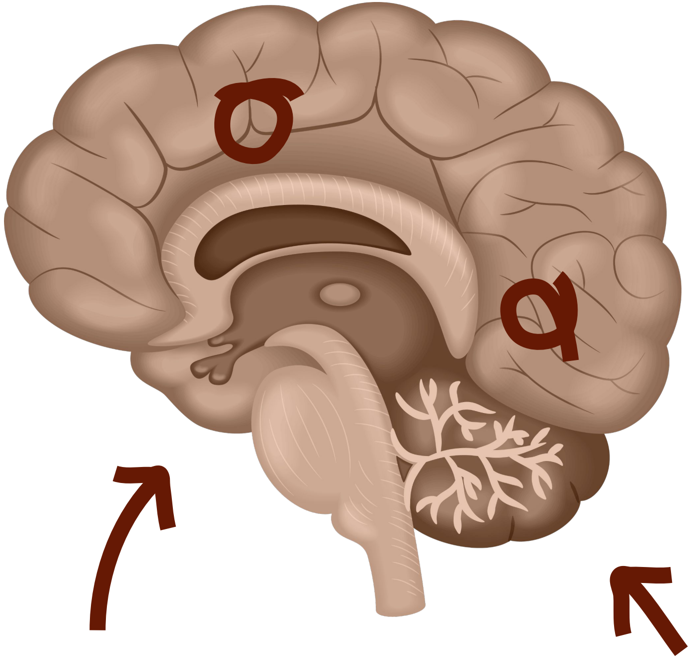
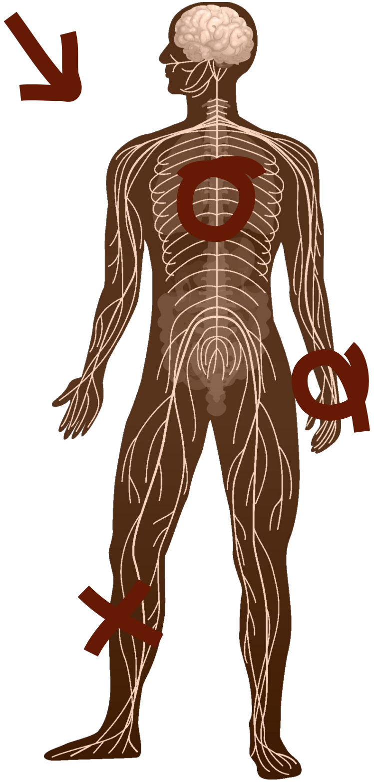
Brain Tumor Classification

Link to All Work:
https://github.com/velocitysix/braintumorML

Using CNN's to classify Brain Tumors.
Prathik Nair





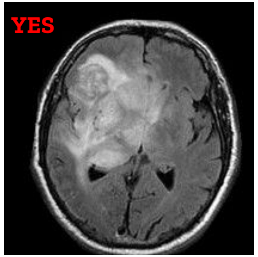
Introduction & Project Scope

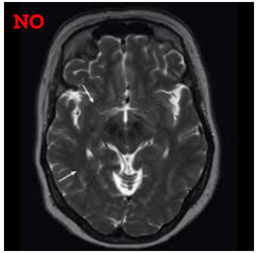
This project aims to tackle the challenge of accurately classifying brain tumors based on medical images & quantitative tumor data, leveraging the power of Convolutional Neural Networks (CNNs).

Our project has the following objectives:

- Employ a Convolutional Neural Network (CNN) model for **image classification.** This model is trained on preprocessed images and utilizes the quantitative features provided for each MRI scan.
- Develop a high-accuracy classifier for brain tumor identification from **quantitative data.** While this project is not involved in medical diagnosis or treatment, the techniques and methodologies applied here could potentially contribute to future research and development

Data Review





Source 1

The brain tumor image dataset and corresponding quantitative features were obtained from the Kaggle dataset 'Brain Tumor' by Jake Shbohaju

Link: https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor).

The dataset consists of five first-order features and eight texture features. The 'Class' column indicates the target level, with 1 denoting the presence of a tumor and 0 representing non-tumor cases. This dataset also contains 3762 MRI images that will be used to test the trained model on unseen data.

- First Order Features
  - Mean
  - Variance
  - Standard Deviation
  - Skewness
  - Kurtosis
- Second Order Features
  - Contrast
  - Energy
  - ASM (Angular second moment)
  - Entropy
  - Homogeneity
  - Dissimilarity
  - Correlation
  - Coarseness

Source 2

Link: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri

The image dataset comprises approximately 3264+ MRI images (.jpg) of brain tumors. Each image is associated with five first-order features and eight second-order texture features, and separated into subdirectories of Glioma (tumor) and NoTumor (No Tumor). While the dataset contains many types of tumors, I selected Glioma and NoTumor to use in the model.

Model Assumptions & Limitations

Assumptions

- The neural network model architecture (MobileNetV2) is suitable for the given brain tumor classification task.
- The selected hyperparameters, such as the number of layers and number of neurons are appropriate for the task.
- The model assumes that the relationship between the input images and the target variable (tumor or non-tumor) can be learned effectively through training.

Limitations

- The model assumes that there are no external factors or external variables that might influence the prediction of brain tumor classification.
- The model may have limitations in handling certain types of brain tumors and may not generalize well to images that significantly differ from the training data.
- The model assumes that the provided features and the extracted features from MobileNetV2 are sufficient for accurate classification.
- Limited image dataset- 3264 MRI images

Hypothesis

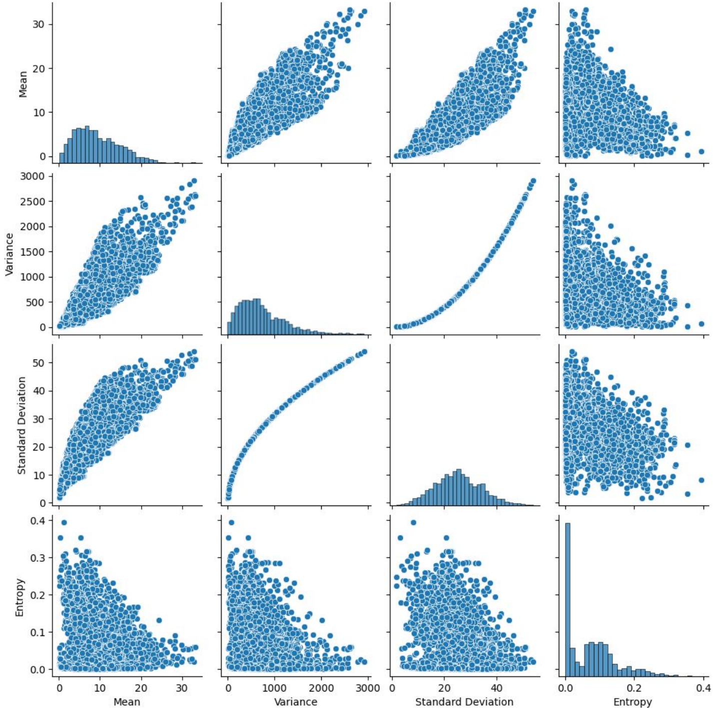
- The hypothesis is that the neural network model, trained on the provided brain tumor dataset, will achieve high accuracy and performance in classifying brain tumor images.
- the model's architecture, training process, and the quality of the dataset will enable it to effectively learn and capture the distinguishing features necessary for accurate classification.

EDA

Dataset Characteristics:

- 44.7% of the images are labeled as class 1 (slight class imbalance)
- Kurtosis, Contrast, and Skewness had high percentile values of Standard Deviation, indicating they have the most outliers.
- Strong Correlation between Variance and STD Deviation (to be expected)

Class	Count	Percentage
Class 0 (No Tumor)	2079	55.26%
Class 1 (Tumor)	1683	44.74%



Model 1 - Binary Classification Neural Network for Brain Tumor Detection

Feature Engineering

- Used SelectKBest with f_classif scoring to select the top features based on their relevance to the target variable.
- The selected features after applying SelectKBest were **Variance, Standard Deviation, Entropy, Skewness, Kurtosis, Contrast, Energy, ASM, Homogeneity, and Dissimilarity.**

Methodology

- Removed the Image (id) and Class column from the dataset. The Class column already indicates which images have tumors. By dropping these columns, we are able to see if the selected features can accurately classify if a tumor is present.
- Utilized a feedforward neural network with multiple dense layers, ReLU activation, and a sigmoid activation function for binary classification.
- Trained the model using the Adam optimizer and binary cross-entropy loss.
- These choices were made to capture complex relations and patterns present in the data.

Model 1 - Findings

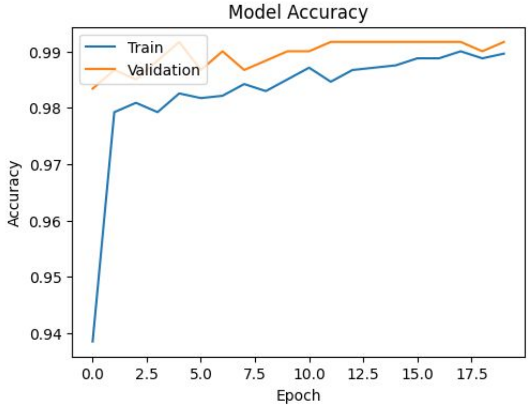
Findings

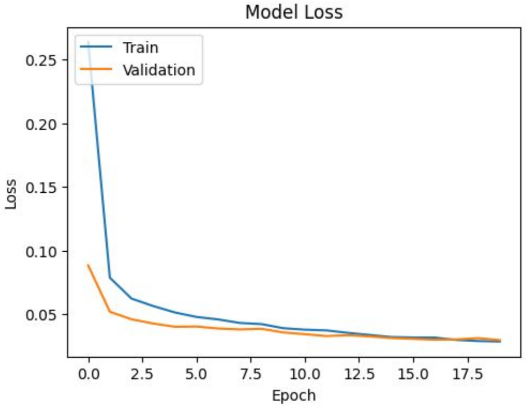
- Train Accuracy: 99.2%
- Test Accuracy: 98.4%
- These findings indicate that the model with the selected features, performs well in predicting the presence/absence of brain tumors on unseen data.

Evaluation

- Accuracy over epochs was plotted to visualize the models learning progress, and to evaluate overfitting and underfitting
- As shown in the plots below. The model is learning well without underfitting or overfitting.
- The metrics listed below indicate that the model performs well in classifying. It achieves high precision, recall, F1, and ROC. This suggests the model is also very accurate and has good balance.

Metric	Value
Precision	.982
Recall	.976
F1	.979
ROC AUC	.981





Model 2 - Image Classification using Transfer Learning with MobileNetV2

Data Processing

- The model extracts and prepocesses images from the input file.
- The images undergo rescaling and augmentation using ImageDataGenerator (keras)
- Corrupted images are removed
- DataGen batches images for training

Methodology

- The base model is built using MobileNetV2- this has pre-loaded weights, and pre-trained layers.
- The base model is extended by adding more layers (average pooling layer, etc)
- Next, the model is compiled with the Adam optimizer and binary cross-entropy loss (final layer uses sigmoid activation function for binary classification)(presence or absence of a tumor)
- Model performance is evaluated during training on the validation set to keep track of the loss and accuracy, which is used to update the model weights. This allows me to assess how well the model is doing during the training process.
- Finally, the model is run on unseen data.

Findings

Validation Test

- The model was evaluated on a set of data that it had not seen during training, in my case, I tested it on validation data and a new tumor dataset.
- The model achieved an accuracy of approximately **92%**, meaning it made correct predictions for 90% of the samples in the dataset it was evaluated on.
- The model achieved a loss of **.248** indicating a good fit

Unseen Data Test

- The model performs reasonably well on unseen data, with an accuracy of over **82%**
- Accuracy was classified by comparing the predicted labels, to the image labels.
- Images with tumors had the word ‘tumor’ in them and images without tumors had the word ‘notumor’ in them.
- Binary classification threshold was set to .5- this can be adjusted to get better sensitivity and perhaps increase accuracy

Future Work

Validation Test

- The model architecture can be optimized with deeper networks and different types of layers
- We can use ensemble methods to learn multiple models and used the combined learning to make better predictions.
- An expanded dataset for training can help improve or uncover biases.
- It was my intention to continue this model to not only tell you if there is a tumor but also where the tumor is in the image; this kind of task requires a different type of model.

Link to All Work: https://github.com/velocitysix/braintumorML